EK287384616US

Docket No. AUS9-2000-0401-US1

METHOD AND SYSTEM FOR BUILDING DYNAMIC FIREWALL RULES, BASED ON CONTENT OF DOWNLOADED DOCUMENTS

BACKGROUND OF THE INVENTION

1. Technical Field:

The present invention is related to construction of firewalls to screen internal computer networks from the Internet. More specifically, the invention screens incoming data based on content.

2. Description of Related Art:

Increased reliance on the Internet in recent years has created a new host of security problems for organizations wishing to exploit this resource. Examples of these problems include the infiltration of computer viruses into internal computer networks, and the downloading of indecent material onto individual workstations. To cope with these problems, organizations have developed several methods for monitoring and controlling the influx of data from the Internet into their internal networks. Each of these approaches has its strengths and weaknesses.

One popular method of filtering incoming Internet data is the use of a firewall, a selective gateway standing between the Internet and an internal computer network. Firewalls can be designed to prevent specific types of data from entering the internal network and have the advantage of providing a centralized point from which administrators can control the influx of data.

A common way of setting up a firewall is to manually insert the address of a particular Internet site into the filtering rules of the firewall. Once a site is listed in the filtering rules, the firewall will automatically prevent electronic documents from that site from entering the internal network. A primary advantage of this method is that specific, objectionable sites can be blocked with certainty. Unfortunately, this approach is also very labor intensive, as it requires administrators to first evaluate the content of an Internet site and then manually add it to the filtering rules. This process diverts administrators from the important job of managing an organization to the mundane job of monitoring Internet access by employees, students, or staff. In addition, the administrator cannot add a site to the filtering list unless he or she knows about it. Considering the size and dynamic nature of the Internet, the administrator is certain to remain several steps behind changes in Internet content.

Another approach is to have the firewall scan the content description language coming in from the Internet and check the tag information within the content description. The tags describe the elements of an electronic document and are used by browser programs to display data properly. If the firewall detects content descriptions that have been added to the filtering rules, the Internet document will automatically be blocked from entering the internal network. Such content filtering reduces the burden on administrators by allowing them to set more general filtering guidelines rather than manually adding individual sites to the filtering rules. A disadvantage of this type of content filtering is the processing burden created by scanning all incoming Internet traffic. Having to scan the content description of all incoming traffic and compare that content to the filtering rules requires considerable processing resources, which must be diverted from other uses.

In addition to firewalls, Internet content may be filtered by using a distributed network, in which Internet content is filtered at the workstation just before it is rendered by the browser. This approach essentially offloads processing tasks from the server onto the client computers, which can substantially degrade performance, especially if the clients are "thin," having little processing capability themselves. This type of setup generally does not work well in a corporate environment because of the performance degradation and logistical problems of having filtering code distributed among several client machines.

Therefore, an Internet filtering method that allows content to be screened at a central firewall, but does not require heavy processing loads or constant monitoring and input from an administrator, is desirable.

SUMMARY OF THE INVENTION

The present invention provides a method for filtering incoming data from an external computer network. This method includes scanning the contents of incoming data for pre-selected keyword(s) and allowing it to pass per standard service rules if its content does not contain the pre-selected keyword(s). If the incoming data does contain pre-selected keywords, it is blocked and added to a "known-block" filtering table. Once added to the filtering table, the site will automatically be blocked in the future without having its contents scanned again for pre-selected keywords.

BRIEF DESCRIPTION OF THE DRAWINGS

The novel features believed characteristic of the invention are set forth in the appended claims. The invention itself, however, as well as a preferred mode of use, further objectives and advantages thereof, will best be understood by reference to the following detailed description of an illustrative embodiment when read in conjunction with the accompanying drawings, wherein:

Figure 1 depicts a block diagram illustrating the relationship between an external computer network, a firewall, and an internal computer network;

Figure 2 depicts a block diagram of a data processing system which may be implemented as a server, in accordance with the present invention; and

Figure 3 depicts a flowchart illustrating the operation of dynamic firewall rules in accordance with the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

With reference now to **Figure 1**, a block diagram illustrating the typical relationship between an external computer network, a firewall, and an internal computer network is depicted. The purpose of the firewall **102** is to act as a selective gateway controlling the flow of information between the external **101** and internal **103** computer networks. The firewall **102** can be designed to prevent users of the internal network **103** from accessing inappropriate material from the external network **101**.

Referring to **Figure 2**, a block diagram of a data processing system which may be implemented as a server is depicted in accordance with the present invention. Data processing system **200** may be a symmetric multiprocessor (SMP) system including a plurality of processors **202** and **204** connected to system bus **206**. Alternatively, a single processor system may be employed. Also connected to system bus **206** is memory controller/cache **208**, which provides an interface to local memory **209**. I/O bus bridge **210** is connected to system bus **206** and provides an interface to I/O bus **212**. Memory controller/cache **208** and I/O bus bridge **210** may be integrated as depicted.

Peripheral component interconnect (PCI) bus bridge **214** connected to I/O bus **212** provides an interface to PCI local bus **216**. A number of modems **218-220** may be connected to PCI bus **216**. Typical PCI bus implementations will support four PCI expansion slots or add-in connectors. Communications links to network computers may be provided through modem **218** and network adapter **220** connected to PCI local bus **216** through add-in boards.

Additional PCI bus bridges **222** and **224** provide interfaces for additional PCI buses **226** and **228**, from which additional modems or network adapters may be supported. In this manner, server **200** allows connections to multiple network computers. A memory mapped graphics adapter **230** and hard disk **232** may also be connected to I/O bus **212** as depicted, either directly or indirectly.

Those of ordinary skill in the art will appreciate that the hardware depicted in **Figure 2** may vary. For example, other peripheral devices, such as optical disk drives and the like, also may be used in addition to or in place of the hardware depicted. The depicted example is not meant to imply architectural limitations with respect to the present invention.

The data processing system depicted in **Figure 2** may be, for example, an IBM RS/6000, a product of International Business Machines Corporation in Armonk, New York, running the Advanced Interactive Executive (AIX) operating system.

With reference to **Figure 3**, a flowchart illustrating the functioning of a dynamic firewall is depicted in accordance with the present invention. In this example, HTML (Hyper Text Markup Language) documents from an external computer network arrive at the firewall due to a request from an internal user (step **301**). The next step is to determine if the HTML document originated from a site already listed in a filtering table of "known-block" sites (step **302**). "Known-block" would refer to any specific site that was already known to contain inappropriate material. If the document is from a "known-block" site, then it is blocked from entering the internal network (step 303).

In one embodiment of the invention, if the document is not from a "known-block" site, the next step is to determine if it is from a "known-safe" site (step 304). "Known-safe" would refer to specific sites that administrators knew to be free of objectionable material or sites that are vital to organizational operations, for example, suppliers. If the document is from a "known-safe" site, it is allowed to pass per standard service rules without having its content scanned (step 305).

If the document is not from a "known-safe" site, the next step is to scan the contents for keywords that have been pre-selected by an administrator (steps 306 and 307). In prior art, when a HTML document is scanned, the contents are determined by looking at the tags in the description language. The tags describe the elements of a document and are used by browser programs to display data properly, but they are not part of the text of the document. The present invention goes a step further than the prior art and actually scans the text fields within the document itself, which provides a more accurate guide to content.

If the document does contain pre-selected keywords, it is blocked (step 308). In addition, once a document has been identified as containing pre-selected keywords, its originating address is automatically added to the filtering table of "known-block" sites (step 309). In the future, any document originating from that address will automatically be blocked by the firewall, without the need to scan its contents again, thus reducing the processing load on the system. Adding the site address to the "known-block" filtering table can be done using PERL or any strong text parsing language. This automatic process reduces the burden on administrators, who do not have to manually add sites to the firewall rules.

The updated filtering table can be added to the firewall instance through periodic refreshes, at intervals ranging, for example, from once a week to once an hour, depending on the needs of the organization in question. The instance is the running copy of the firewall loaded into memory, and must be refreshed to incorporate the new rules. The system can refresh the firewall instance by means of a timed job on a Windows platform, a cron job on a UNIX platform, or the equivalent on any other operating system which would be used for a firewall host.

If the HTML document does not contain any pre-selected keywords, it will pass into the internal network per standard service rules (step 310).

It should be pointed out that in addition to HTML documents, the present invention also applies to computer documents written in other languages, such as, for example, XML (Extensible Markup Language).

It is important to note that while the present invention has been described in the context of a fully functioning data processing system, those of ordinary skill in the art will appreciate that the processes of the present invention are capable of being distributed in the form of a computer readable medium of instructions and a variety of forms and that the present invention applies equally regardless of the particular type of signal bearing media actually used to carry out the distribution. Examples of computer readable media include recordable-type media, such as a floppy disk, a hard disk drive, a RAM, CD-ROMs, DVD-ROMs, and transmission-type media, such as digital and analog communications links, wired or wireless communications links using transmission forms, such as, for example, radio frequency and light wave transmissions. The computer readable media may take the form of coded formats that are decoded for actual use in a particular data processing system.

The description of the present invention has been presented for purposes of illustration and description, and is not intended to be exhaustive or limited to the invention in the form disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art. The embodiment was chosen and described in order to best explain the principles of the invention, the practical application, and to enable others of ordinary skill in the art to understand the invention for various embodiments with various modifications as are suited to the particular use contemplated.